(12) EUROPEAN PATENT APPLICATION

(72) Inventors:
• Baumgarte, Frank
Watchung, New Jersey 07069 (US)
• Kroon, Petter
Green Brook, New Jersey 08812 (US)

(74) Representative: Williams, David John et al
Page White & Farrer,
54 Doughty Street
London WC1N 2LS (GB)

(54) Hybrid multi-channel/cue coding/decoding of audio signals

(57) Part of the spectrum of two or more input signals is encoded using conventional coding techniques, while encoding the rest of the spectrum using binaural cue coding (BCC). In BCC coding, spectral components of the input signals are downmixed and BCC parameters (e.g., inter-channel level and/or time differences) are generated. In a stereo implementation, after converting the left and right channels to the frequency domain, pairs of left- and right-channel spectral components are downmixed to mono. The mono components are then converted back to the time domain, along with those left- and right-channel spectral components that were not downmixed, to form hybrid stereo signals, which can then be encoded using conventional coding techniques. For playback, the encoded bitstream is decoded using conventional decoding techniques. BCC synthesis techniques may then apply the BCC parameters to synthesize an auditory scene based on the mono components as well as the unmixed stereo components.

FIG. 1
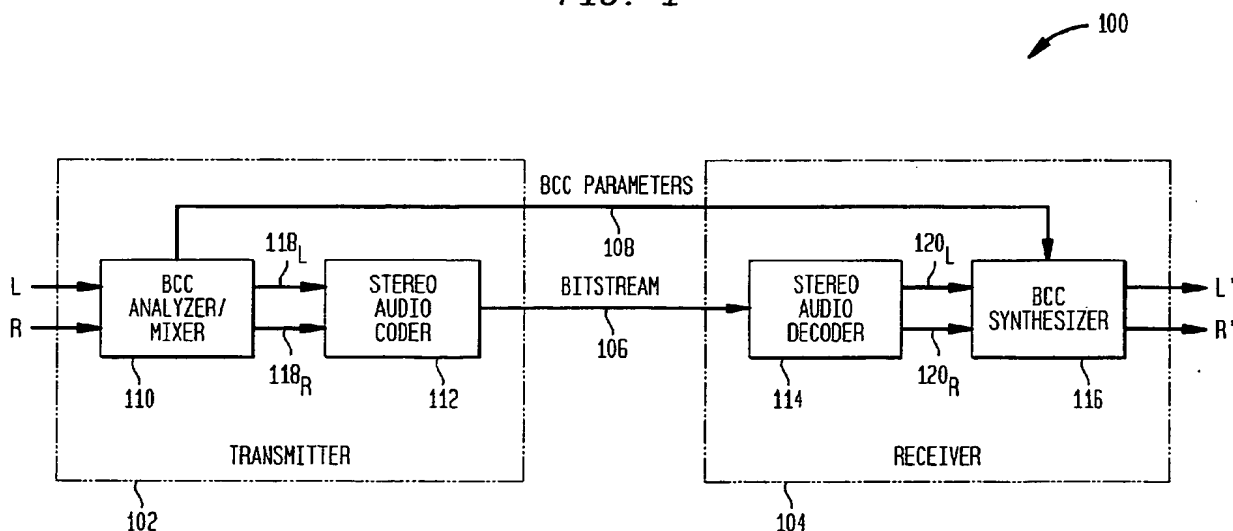
EP 1 376 538 A1

Description

[0001] The present invention relates to the encoding of audio signals and the subsequent decoding of the encoded audio signals to generate auditory scene during playback.

[0002] In conventional stereo audio coding, the sum and the difference of the left and right audio channels of the stereo input signal are formed and then individually coded, e.g., using adaptive differential pulse code modulation (ADPCM) or some other suitable audio coding algorithm. to form an encoded audio bitstream. The corresponding conventional stereo audio decoding involves reversing the (ADPCM) coding algorithm to recover decoded sum and difference signals, from which left and right audio channels of a decoded stereo output signal are generated.

Although such conventional stereo audio coding/decoding (codec) techniques can produce an auditory scene during playback that accurately reflects the fidelity of the stereo input signal, the amount of data required for the corresponding encoded audio bitstream may be prohibitively large for some applications involving limited storage space and/or transmission bandwidth.

[0003] US Patent Serial number 09/848,877, 10/045,458 and 10/155,437 describe audio codec techniques that can produce smaller encoded audio bitstreams for the same or substantially similar levels of playback fidelity as those associated with conventional stereo audio codecs. In particular, these patent applications are related to an audio coding technique referred to as binaural cue coding (BCC).

[0004] When BCC coding is applied to stereo audio, the left and right channels of the stereo input signal are downmixed (e.g., by summing) to a single mono signal, which is then encoded using a suitable conventional audio coding algorithm such as ADPCM. In addition, the left and right channels are analyzed to generate a stream of BCC parameters. In one implementation, for each audio frame (e.g., 20 msec), the BCC parameters include auditory spatial cues such as an inter-channel or inter-aural level difference (ILD) value and an inter-channel or inter-aural time difference (ITD) value between the left and right channels for each of a plurality of different frequency bands in the stereo input signal. Since the corresponding encoded audio data might include only an encoded mono signal and a stream of BCC parameters, the amount of encoded data may be considerably smaller (e.g., 50-80%) than that for a corresponding encoded audio bitstream generated using conventional stereo audio coding, such as that described previously.

[0005] The corresponding BCC decoding involves reversing the (e.g., ADPCM) coding algorithm to recover a decoded mono signal. Stereo audio synthesis techniques are then applied to the decoded mono signal using the BCC parameters to generate left and right channels of a decoded stereo audio signal for playback. Al-

though typically lower than that achieved using conventional stereo audio codecs, the fidelity of an auditory scene generated using BCC coding and decoding may be acceptable for many applications, while typically using lower bandwidth.

[0006] Embodiments of the present invention are related to a hybrid audio codec technique in which conventional audio coding is applied to certain frequency bands of the input audio signals, while BCC coding is applied to other frequency bands of the input audio signals. In one possible stereo implementation, signal spectral components whose frequencies above a specified threshold frequency (e.g., 1.5 kHz) are coded using BCC coding, while lower-frequency components are coded using conventional stereo coding. As a result, even higher fidelity playback can be achieved than using only BCC coding, while still reducing the total amount of encoded data compared to conventional stereo coding.

[0007] According to one embodiment, the present invention is a method for encoding N input audio signals, where $N > 1$. Each of the N input audio signals is converted into a plurality of spectral components in a frequency domain. For each of one or more, but not all, of the spectral components, the spectral components corresponding to the N input audio signals are downmixed to generate a downmixed spectral component, leaving one or more of the spectral components for each of the N input audio signals unmixed. An encoded audio bitstream is generated based on the one or more downmixed spectral components and one or more unmixed spectral components.

[0008] Preferably, the one or more auditory spatial parameters include one or more of an inter-channel level difference and an inter-channel time difference.

[0009] Preferably, the one or more downmixed spectral components have frequencies above a specified threshold frequency, and the one or more unmixed spectral components have frequencies below the specified threshold frequency.

[0010] Preferably, the specified threshold frequency varies dynamically over time.

[0011] Preferably the specified threshold frequency varies as a function of bit rate.

[0012] Preferably the one or more down mixed spectral components have spectral energies below a specified threshold energy, and the one or more unmixed spectral components have spectral energies above the specified threshold energy.

[0013] According to another embodiment, the present invention is an encoded audio bitstream generated by performing the previously recited method.

[0014] According to another embodiment, the present invention is an apparatus for processing N input audio signals, where N>1, for encoding. One or more transform are configured to convert each of the N input audio signals into a plurality of spectral components in a frequency domain. A downmixer is configured, for each of one or more, but not all, of the spectral components, to

downmix the spectral components corresponding to the N input audio signals to generate a downmixed spectral component, leaving one or more of the spectral components for each of the N input audio signals unmixed.

[0015] Preferably, the apparatus further comprises one or more inverse transforms configured to convert the one or more downmixed spectral components and the one or more unmixed spectral components into N hybrid audio signals.

[0016] Preferably, the apparatus further comprises an audio coder configured to generate an encoded audio bitstream based on the one or more downmixed spectral components and the one or more unmixed spectral components.

[0017] Preferably, the apparatus further comprises a generator configured to generate one or more auditory spatial parameters for the one or more downmixed spectral components.

[0018] Preferably, N=2, the two input audio signals correspond to left and right input audio signals of a stereo input audio signal, each downmixed spectral component is a mono spectral component, and a stereo audio coder can generate an encoded audio bitstream based on the one or more downmixed spectral components and the one or more unmixed spectral components.

[0019] Preferably, the one or more downmixed spectral components have frequencies above a specified threshold frequency, and the one or more unmixed spectral components have frequencies below the specified threshold frequency.

[0020] According to another embodiment, the present invention is a method for decoding an encoded audio bitstream. The encoded audio bitstream is decoded to generate a plurality of spectral components in a frequency domain, wherein one or more sets of the spectral components correspond to downmixed spectral components, and one or more sets of the spectral components correspond to unmixed spectral components. For each set of the downmixed spectral components, one or more auditory spatial parameters are applied to generate a synthesixed spectral component. The synthesized spectral components and the unmixed spectral components are converted into N decoded audio signals in a time domain. where N>1.

[0021] Preferably the one or more down mixed spectral components have frequencies above a specified threshold frequency, and the one or more unmixed spectral components have frequencies below the specified threshold frequency.

[0022] According to another embodiment. the present invention is an apparatus for decoding an encoded audio bitstream. An audio decoder is configured to decode the encoded audio bitstream to generate a plurality of spectral components in a frequency domain. wherein one or more sets of the spectral components correspond to downmixed spectral components, and one or more sets of the spectral components correspond to unmixed spectral components. A synthesizer is config-

ured, for each set of the downmixed spectral components, to apply one or more auditory spatial parameters to generate a synthesized spectral component. One or more inverse transforms are configured to convert the synthesized spectral components and the unmixed spectral components into N decoded audio signals in a time domain, where N>1.

[0023] Preferably the audio decoder is configured to decode the encoded audio bitstream to generate N hybrid audio signals and further comprising one or more transforms configured to convert each of the N hybrid audio signals into the plurality of spectral components in the frequency domain.

[0024] Preferably, N=2, the encoded audio bitstream is decoded using a stereo audio decoder, the two hybrid audio signals correspond to left and right hybrid audio signals of a hybrid stereo audio signal, and each downmixed spectral component is a mono spectral component.

[0025] Preferably, the one or more downmixed spectral components have frequencies above a specified threshold frequency and the one or more unmixed spectral components have frequencies below the specified threshold frequency.

[0026] Other aspects, features, and advantages of the present invention will become more fully apparent from the following detailed description, the appended claims, and the accompanying drawings in which:

Fig 1 shows a block diagram of a hybrid audio system, according to one embodiment of the present invention:

Fig 2 shows a block diagram of the processing implemented by the BCC analyser/mixer of Fig. 1, according to one embodiment of the present invention; and

Fig 3 shows a block diagram of the processing implemented by the BCC synthesizer of Fig 1, according to one embodiment of the present invention.

[0027] Fig. 1 shows a block diagram of a hybrid audio system 100, according to one embodiment of the present invention. Audio system 100 comprises transmitter 102 and receiver 104. Transmitter 102 receives the left (L) and right (R) channels of an input stereo audio signal and generates an encoded audio bitstream 106 and a corresponding stream 108 of BCC parameters, which, depending on the implementation, may or may not be explicitly encoded into bitstream 106. Fig. 1 shows BCC parameter stream 108 being transmitted out-of-band from transmitter 102 to receiver 104. In either case, receiver 104 receives the data generated by transmitter 102, decodes encoded audio bitstream 106, and applies the BCC parameters in stream 108 to generate the left (L') and right (R') channels of a decoded stereo audio signal.

[0028] More particularly, transmitter 102 comprises BCC analyzer/mixer 110 and stereo audio coder 112,

while receiver **104** comprises stereo audio decoder **114** and BCC synthesizer **116**.

[0029] In transmitter **102**, BCC analyzer/mixer **110** converts the left (L) and right (R) audio signals into the frequency domain. For spectral components above a specified threshold frequency, BCC analyzer/mixer **110** generates BCC parameters for stream **108** using the BCC techniques described in the '877, '458, and '437 applications. BCC analyzer/mixer **110** also downmixes those high-frequency components to mono. Copies of the high-frequency mono component are then converted back to the time domain in combination with the low-frequency "unmixed" left and right components (i.e., the unprocessed frequency-domain components below the specified threshold frequency), respectively, to form hybrid left and right signals **118**. Stereo audio coder **112** applies conventional stereo coding to these hybrid left and right signals to generate encoded audio bitstream **106**.

[0030] Fig. 2 shows a block diagram of the processing implemented by BCC analyzer/mixer **110** of Fig. 1, according to one embodiment of the present invention. Fast Fourier transform (FFT) 202_L converts the left audio signal L into a plurality of left-channel spectral components **204** in the frequency domain. Similarly, FFT 202_R converts the right audio signal R into a plurality of right-channel spectral components **206** in the frequency domain. The one or more left-channel components 204_{HI} and the corresponding one or more right-channel components 206_{HI} whose frequencies are above the specified threshold frequency are applied to both downmixer **208** and BCC parameter generator **216**.

[0031] Downmixer **208** combines each high-frequency left-channel component 204_{HI} with its corresponding high-frequency right-channel component 206_{HI} to form a high-frequency mono component 210_{HI}. The processing performed by downmixer **208** to generate the mono components may vary from implementation to implementation. In one possible implementation, downmixer **208** simply averages the corresponding left-and right-channel components. In another possible implementation, downmixer **208** implements the downmixing technique described in the 'xxx application. Those skilled in the art will appreciate that other suitable downmixing algorithms are possible.

[0032] Replicator **212** generates two copies of each high-frequency mono component 210_{HI} for application to left and right inverse FFTs (IFFTs) 214_L and 214_R, respectively. IFFTs 214_L and 214_R also receive the low-frequency left and right components 204_{LO} and 206_{LO}, respectively, from FFTs 202_L and 202_R. IFFTs 214_L and 214_R convert their respective sets of components back to the time domain to generate the left and right hybrid signals 118_L and 118_R, respectively. The resulting two-channel signal contains identical frequency components within spectral regions that were converted to mono, with the remaining parts being identical to the input signals L and R. As a result, stereo audio coder **112** will

typically generate an encoded audio bitstream that has fewer bits than if it were to encode the original input stereo audio signal (L and R).

[0033] BCC parameter generator **216** analyzes the high-frequency left and right components 204_{HI} and 206_{HI} to generate BCC parameters for stream **108** of Fig. 1 for each frequency band above the specified threshold frequency.

[0034] Referring again to Fig. 1, in receiver **104**, stereo audio decoder **114** applies a conventional stereo decoding algorithm (e.g., to reverse the coding implemented by coder **112**) to recover hybrid decoded left and right signals **120**. BCC synthesizer **116** applies BCC synthesis techniques to the high-frequency portions of channels 120 to synthesize the high-frequency portions of the decoded left (L') and right (R') channels. In particular, BCC synthesizer **116** converts the hybrid channels **120** to the frequency domain, applies the BCC parameters to the high-frequency components to synthesize high-frequency left and right components using the BCC techniques described in the '877, '458, and '437 applications, and then reconverts the resulting synthesized high-frequency components and corresponding decoded low-frequency components to the time domain.

[0035] Fig. 3 shows a block diagram of the processing implemented by BCC synthesizer **116** of Fig. 1, according to one embodiment of the present invention. FFT 302_L converts hybrid left audio signal 120_L from stereo audio decoder **114** into a plurality of left-channel spectral components **304** in the frequency domain. Similarly, FFT 302_R converts hybrid right audio signal 120_R from decoder **114** into a plurality of right-channel spectral components **306** in the frequency domain. The one or more left-channel components 304_{HI} and the corresponding one or more right-channel components 306_{HI} whose frequencies are above the specified threshold frequency are applied to mono signal generator **308**.

[0036] Mono signal generator 308 generates a high-frequency mono component for each high-frequency left-channel component 304_{HI} and its corresponding high-frequency right-channel component 306_{HI}. Ideally, since replicator **212** of Fig. 2 generated identical copies of each high-frequency mono component 210_{HI}, each high-frequency left-channel component 304_{HI} should be identical to its corresponding high-frequency right-channel component 306_{HI}. As such, mono signal generator **308** could simply select either the left channel or the right channel to "generate" the one or more high-frequency mono components 310_{HI}. Alternatively, mono signal generator **308** could simply average or perform some other suitable downmixing algorithm, including the algorithm described in the 'xxx application, to generate each mono component 310_{HI}, in order to account for any real-world differences that may exist between the left and right high-frequency component 304_{HI} and 306_{HI}.

[0037] In any case, BCC stereo synthesizer **312** applies BCC processing to generate a high-frequency left-channel component 314_{HI} and a high-frequency right-

channel component 316_{HI} for each high-frequency mono component 310_{HI}. The high-frequency left- and right-channel components 314_{HI} and 316_{HI} are applied to left and right IFFTs 318_L and 318_R, respectively. IFFTs 214_L and 214_R also receive the low-frequency left and right components 304_{LO} and 306_{LO}, respectively, from FFTs 302_L and 302_R. IFFTS 318_L and 318_R convert their respective sets of components back to the time domain to generate left and right channels L' and R', respectively, of the decoded stereo signal of Fig. 1.

[0038] A natural cross-over frequency from the "true" stereo part to the BCC-generated stereo part is 1.5 kHz. Above that frequency the human auditory system does not substantially evaluate inter-aural phase differences for sound localization. Thus, the human auditory system is less sensitive to inter-channel phase errors introduced by BCC processing in that range. Moreover, the most salient auditory localization cues are usually derived from low-frequency components, unless the audio signal has dominant spectral energy at higher frequencies.

[0039] The present invention can also be implemented using a hybrid transmitter such as transmitter 102 of Fig. 1, but a receiver that does not perform any BCC processing. In this case, BCC synthesizer 116 of Fig. 1 may be omitted from receiver 104, and the resulting receiver can ignore BCC parameter stream 108 during decoding processing. Legacy receivers that contain only a conventional audio decoder fall into that category. Such a receiver would not provide BCC spatialization of the auditory image for spectral parts of the decoded audio signals that are based on mono components. However, there is still a remaining stereo effect created by those parts of the spectrum that are preserved as stereo. This stereo effect by itself provides a mechanism for bit-rate reduction as compared to the transmission of the full-bandwidth stereo. Explicitly, mixing parts of the spectrum of the audio input signal to mono reduces the bit rate of a conventional audio coder. The spatial image degradation should be tolerable, if the mono part of the spectrum is limited to frequencies above about 1 kHz.

[0040] For some applications, BCC processing may be intentionally limited to transmit only inter-channel level differences as the BCC parameters (i.e., and not any inter-channel time differences). For headphone playback, inter-channel time differences are important for creating a natural spatial image, especially at frequencies below 1,5 kHz. By keeping the stereo signal up to a limit of about 1.5 kHz, the spatial cues in that frequency are available at the receiver and greatly improve the listening experience with headphones.

[0041] Transmitting a small spectral bandwidth as a stereo signal does not necessarily increase the bit rate of the audio coder dramatically compared to applying BCC processing to the full spectral range. The audio coder can still take full advantage of those parts of the spectrum that are mono by using, e.g., sum/difference coding. The data rate for the BCC parameters can be reduced, since no parameters need to be transmitted for the spectral part that is kept stereo.

[0042] The application of BCC processing to spectral regions can be made adaptive such that an optimum quality/bit-rate tradeoff is achieved. For instance, BCC processing could be switched off for very critical material, or it could be applied to the full spectrum for non-critical material. The spectral region where BCC processing is applied can be controlled, for instance, by one parameter per frame that indicates the upper frequency bound up to which the stereo signal is kept for encoding. In addition, the threshold frequency between stereo and BCC coding could dynamically change based on the number of bits that would actually be used to code different spectral regions of the audio data by the different techniques.

[0043] The audio quality range covered by the hybrid codec scheme in Fig. 1 reaches transparent quality when the spectral region of BCC processing has zero bandwidth. With continuously increasing bandwidth for BCC processing, a gradual quality transition from traditional stereo audio coding to the original full-bandwidth BCC coding scheme of the '877, '458, and '437 applications is possible. Therefore, the quality range of the present invention extends to both quality ranges: that of the original BCC scheme and that of the traditional audio coding scheme.

[0044] Moreover, the hybrid coding scheme is inherently bit-rate scalable. In terms of the coder structure, such a scheme is also referred to as "layered coding." This feature can be used for instance to reduce the bit rate of a given bitstream to accommodate for channels with lower capacity. For such purposes, the BCC parameters can be removed from the bitstream. In that case, a receiver is still able to decode an audio signal with a reduced stereo image, as described above for the legacy decoder. A further step for reducing the bit rate is meaningful, if the stereo audio coder uses sum/difference coding. It is possible to isolate the difference signal information in the bitstream and remove it. In this case, the receiver will decode only the sum signal, which is a monophonic audio signal.

[0045] The different "layers" (e.g., sum, difference, and BCC information) also provide a natural division of the bitstream for unequal error protection for lossy channels. For such applications, the sum signal would get the highest protection and the BCC information would get the lowest protection. If the channel temporarily has a high error rate, then the mono sum signal might still be recoverable, while the difference signal and BCC information might be lost. Such a scheme avoids more audibly annoying frame concealment mechanisms.

[0046] Although the present invention has been described in the context of applications in which BCC processing is applied to all and only frequency bands above a specified threshold frequency, the present invention is not so limited. In general, for the hybrid processing of the present invention, BCC processing

can be applied to any one or more -- but less than all -- frequency bands, whether they are contiguous or not, and independent of any threshold frequency.

[0047] For example, in one possible implementation. BCC processing is applied to only those frequency bands with energy levels below a specified threshold energy, while conventional stereo encoding is applied to the remaining frequency bands. In this way, conventional stereo encoding optimizes fidelity for the "important" (i.e., high spectral energy) frequency bands, while BCC processing optimizes bandwidth for the less-important (i.e., low spectral energy) frequency bands.

[0048] Although the present invention has been described in the context of encoding and decoding a stereo audio signal, the present invention can also be applied to multi-channel applications having more than two input and output channels. Furthermore, the present invention can be applied to applications in which the number of input channels differs from (either higher or lower than) the number of output channels.

[0049] Although the present invention has been described in the context of receivers that apply the BCC techniques of the '877, '458, and '437 applications to synthesize auditory scenes, the present invention can also be implemented in the context of receivers that apply other techniques for synthesizing auditory scenes that do not necessarily rely on the techniques of the '877, '458, and '437 applications.

[0050] Although the present invention has been described in the context of a real-time system in which the generated data are transmitted immediately from the transmitter to the receiver for real-time decoding and playback, the invention is not so limited. For example, the data generated by the transmitter may be stored in computer memory or other electronic storage medium for subsequent, non-real-time playback by one or more receivers.

[0051] Although the present invention has been described in the context of embodiments having an audio coder (e.g., stereo coder 112 of Fig. 1) that encodes hybrid signals in the time domain to generate an encoded audio bitstream and an audio decoder (e.g., stereo decoder 114) that decodes the encoded audio bitstream to recover decoded hybrid signals in the time domain, the present invention is not so limited. Those skilled in the art will understand that the present invention can be implemented in the context of embodiments that code and decode audio data in the frequency domain. For example, the embodiment of Figs. 1-3 can be modified to replace stereo audio coder 112 and stereo audio decoder 114 with audio codecs that encode and decode, respectively, audio data in the frequency domain. In that case, BCC analyzer/mixer 110 of Fig. 2 can be modified to eliminate replicator 212 and IFFTs 214, and BCC synthesizer 116 of Fig. 3 can be modified to eliminate FFTs 302 and mono signal generator 308. In that case, downmixed (i.e., mono) spectral components 210_{HI} generated by downmixer 208 and unmixed spectral compo-

nents 204_{LO} and 206_{LO} are passed directly to the frequency-domain audio coder in the transmitter. Similarly, the corresponding downmixed (i.e., mono) and unmixed spectral components recovered by the frequency-domain audio decoder in the receiver are passed directly to BCC stereo synthesizer 312 and IFFTs 318, respectively.

[0052] The present invention may be implemented as circuit-based processes, including possible implementation on a single integrated circuit. As would be apparent to one skilled in the art, various functions of circuit elements may also be implemented as processing steps in a software program. Such software may be employed in, for example, a digital signal processor, micro-controller, or general-purpose computer.

[0053] The present invention can be embodied in the form of methods and apparatuses for practicing those methods. The present invention can also be embodied in the form of program code embodied in tangible media, such as floppy diskettes, CD-ROMs, hard drives, or any other machine-readable storage medium, wherein, when the program code is loaded into and executed by a machine, such as a computer, the machine becomes an apparatus for practicing the invention. The present invention can also be embodied in the form of program code, for example, whether stored in a storage medium, loaded into and/or executed by a machine, or transmitted over some transmission medium or carrier, such as over electrical wiring or cabling, through fiber optics, or via electromagnetic radiation, wherein, when the program code is loaded into and executed by a machine, such as a computer, the machine becomes an apparatus for practicing the invention. When implemented on a general-purpose processor, the program code segments combine with the processor to provide a unique device that operates analogously to specific logic circuits.

[0054] It will be further understood that various changes in the details, materials, and arrangements of the parts which have been described and illustrated in order to explain the nature of this invention may be made by those skilled in the art without departing from the scope of the invention as expressed in the following claims.

Claims

1. A method for encoding N input audio signals, $N>1$, comprising the steps of:

(a) converting each of the N input audio signals into a plurality of spectral components in a frequency domain;
(b) for each of one or more, but not all, of the spectral components, downmixing the spectral components corresponding to the N input audio signals to generate a downmixed spectral component, leaving one or more of the spectral

components for each of the N input audio signals unmixed; and

(c) generating an encoded audio bitstream based on the one or more downmixed spectral components and one or more unmixed spectral components.

2. A method of generating an encoded audio bitstream, comprising the steps of:

(a) converting each of N input audio signals, $N>1$ into a plurality of spectral components in a frequency domain;
(b) for each of one or more, but not all, of the spectral components, downmixing the spectral components corresponding to the N input audio signals to generate a downmixed spectral component, leaving one or more of the spectral components for each of the N input audio signals unmixed; and
(c) generating the encoded audio bitstream based on the one or more downmixed spectral components and one or more unmixed spectral components.

3. The method as claimed in claim 1 or 2, wherein step (c) comprises the steps of:

(1) converting the one or more downmixed spectral components and the one or more unmixed spectral components into N hybrid audio signals in a time domain; and
(2) applying an audio coding algorithm to the N hybrid audio signals to generate the encoded audio bitstream.

4. The method as claimed in any of claims 1 to 3, wherein step (b) further comprises the step of generating one or more auditory spatial parameters for the one or more downmixed spectral components.

5. The method as claimed in any of claims 1 to 4, wherein: N=2; the two input audio signals correspond to left and right input audio signals of a stereo input audio signal; each downmixed spectral component is a mono spectral component; and the encoded audio bitstream is generated using a stereo audio coder.

6. An apparatus for processing N input audio signals, $N>1$ for encoding, comprising:

(a) one or more transforms configured to convert each of the N input audio signals into a plurality of spectral components in a frequency domain; and
(b) a downmixer configured, for each of one or more, but not all, of the spectral components,

to downmix the spectral components corresponding to the N input audio signals to generate a downmixed spectral component, leaving one or more of the spectral components for each of the N input audio signals unmixed.

7. A method for decoding an encoded audio bitstream, comprising the steps of:

(a) decoding the encoded audio bitstream to generate a plurality of spectral components in a frequency domain, wherein:

one or more sets of the spectral components correspond to downmixed spectral components; and
one or more sets of the spectral components correspond to unmixed spectral components;

(b) for each set of the downmixed spectral components, applying one or more auditory spatial parameters to generate a synthesized spectral component; and
(c) converting the synthesized spectral components and the unmixed spectral components into N decoded audio signals in a time domain, $N>1$.

8. The method as claimed in claim 7, wherein step (a) comprises the steps of:

(1) decoding the encoded audio bitstream to recover N hybrid audio signals; and
(2) converting each of the N hybrid audio signals into the plurality of spectral components in the frequency domain.

9. The method as claimed in claim 8, wherein: N=2; the encoded audio bitstream is decoded using a stereo audio decoder; the two hybrid audio signals correspond to left and right hybrid audio signals of a hybrid stereo audio signal; and each downmixed spectral component is a mono spectral component.

10. An apparatus for decoding an encoded audio bitstream, comprising:

(a) an audio decoder configured to decode the encoded audio bitstream to generate a plurality of spectral components in a frequency domain, wherein:

one or more sets of the spectral components correspond to downmixed spectral components; and
one or more sets of the spectral components correspond to unmixed spectral

components;

(b) a synthesizer configured, for each set of the downmixed spectral components, to apply one or more auditory spatial parameters to generate a synthesized spectral component; and
(c) one or more inverse transforms configured to convert the synthesized spectral components and the unmixed spectral components into N decoded audio signals in a time domain, $N>1$.
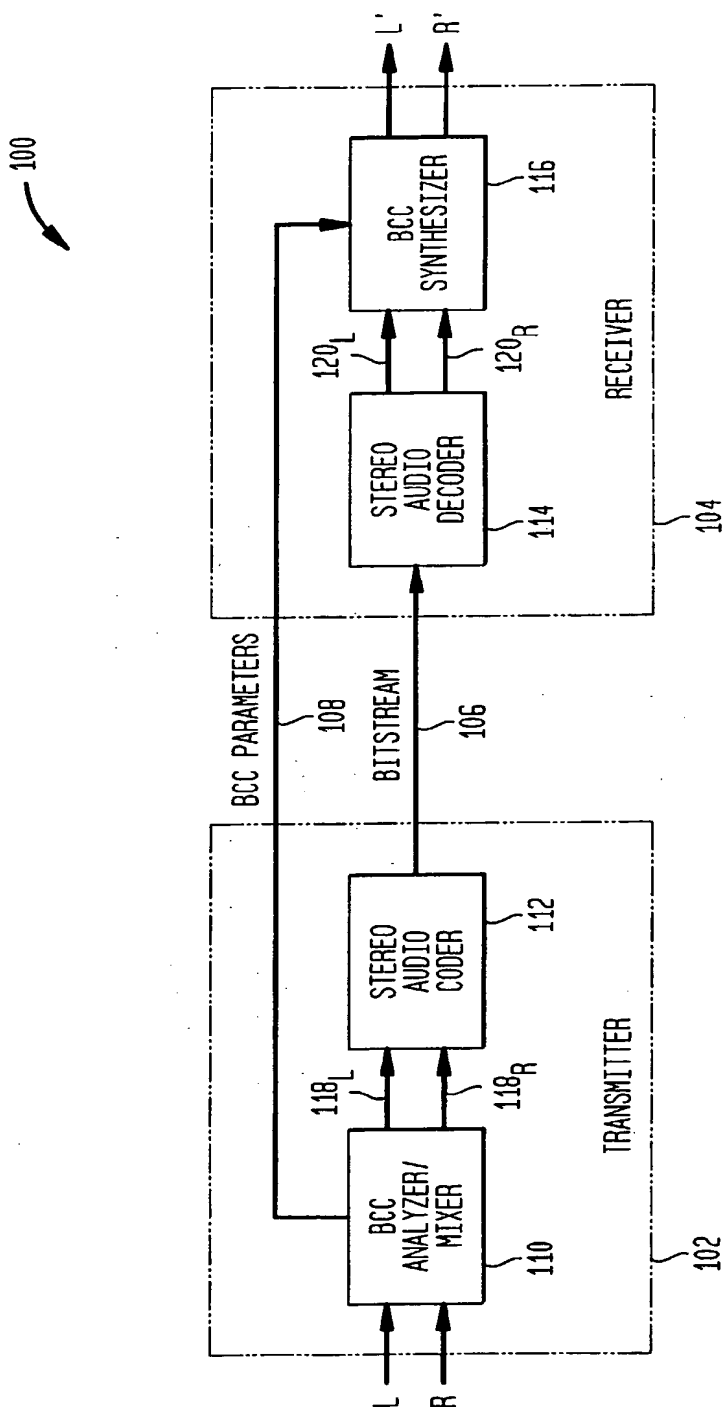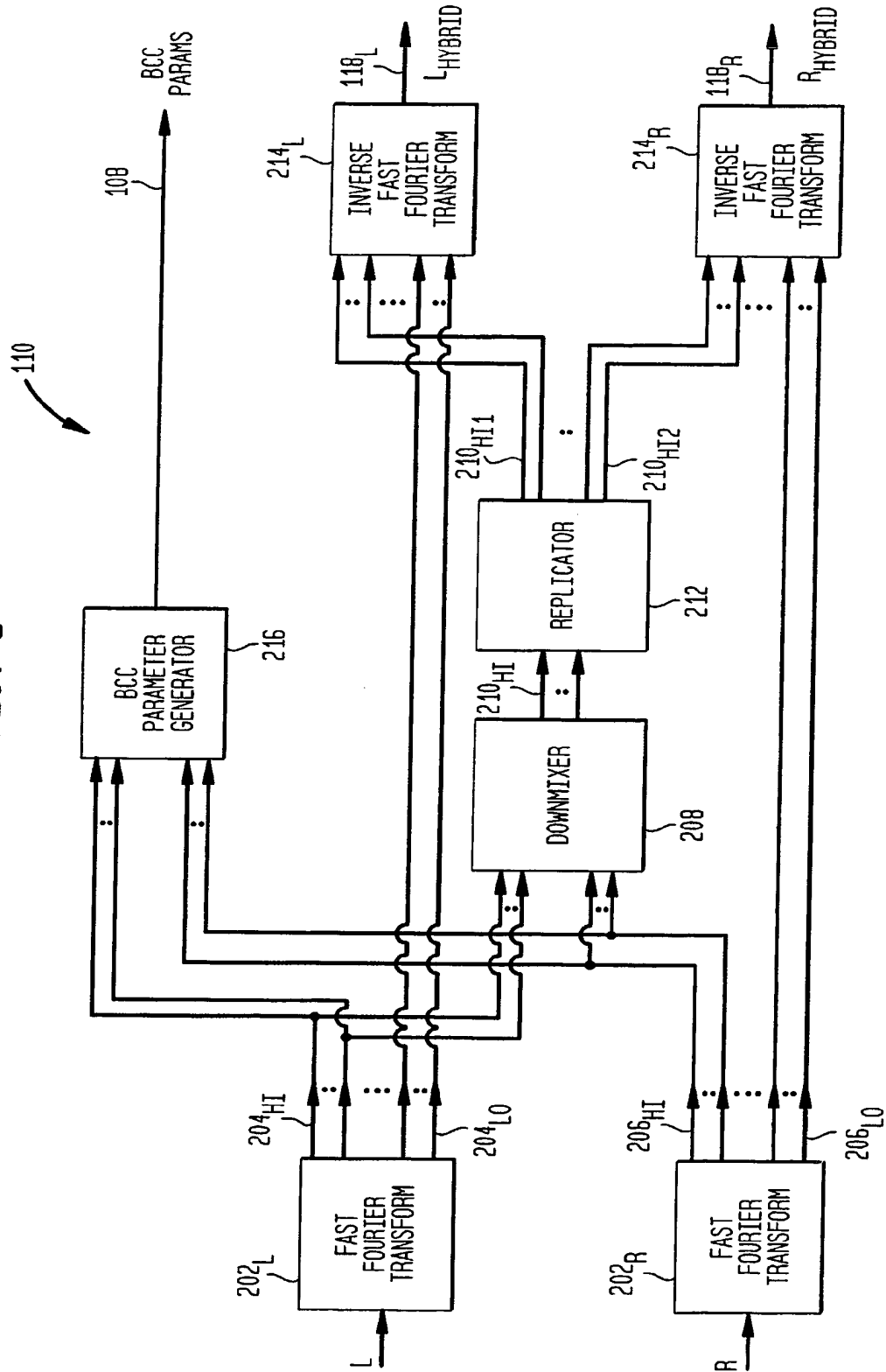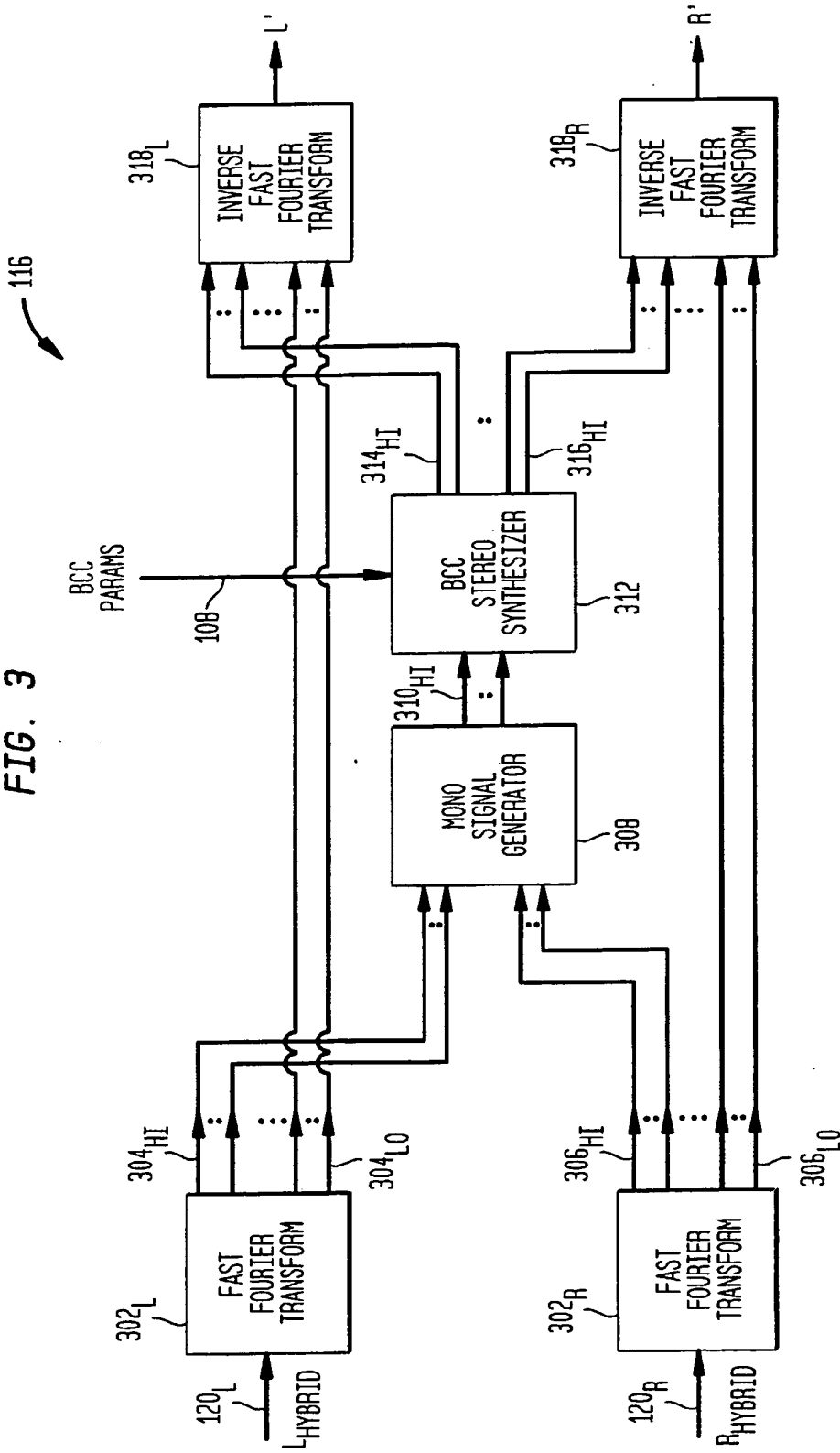
5

10

15

20

25

30

35

40

45

50

55

FIG. 1

FIG. 2

FIG. 3

European Patent
Office

EUROPEAN SEARCH REPORT

Application Number

EP 03 25 3970

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
X	US 5 583 962 A (DOLBY RAY M ET AL) 10 December 1996 (1996-12-10) abstract * figures 6-8 * * column 13, line 2 - line 5 * * column 8, line 25 - line 31 * * column 8, line 65 - column 9, line 38 * * column 23, line 34 - column 24, line 28 * * column 26, line 15 - line 28 * * column 27, line 38 - line 43 * * column 27, line 65 - column 29, line 9 * * column 28, line 2 - line 5 *	1,2,4,6, 7,10	G10L19/02 H04S1/00 H04S5/00
Y	abstract line 1	5	
Y	US 5 703 999 A (EBERLEIN ERNST ET AL) 30 December 1997 (1997-12-30) * column 5, line 14 - line 15; figure 1 *	5	
A	FALLER C ET AL: "Efficient representation of spatial audio using perceptual parametrization" IEEE WORKSHOP ON APPLICATIONS OF SIGNAL PROCESSING TO AUDIO AND ACOUSTICS, 21 October 2001 (2001-10-21), pages 199-202, XP010566909 New Paltz,NY abstract, sections 1-2	1-10	TECHNICAL FIELDS SEARCHED (Int.Cl.7) G10L H04S

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
MUNICH	18 September 2003	Dobler, E

EPO FORM 1503 03.82 (P04C01)

12

ANNEX TO THE EUROPEAN SEARCH REPORT
ON EUROPEAN PATENT APPLICATION NO.

EP 03 25 3970

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

18-09-2003

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
US 5583962	A	10-12-1996	US	6021386 A	01-02-2000
			US	5633981 A	27-05-1997
			US	5632005 A	20-05-1997
			US	5909664 A	01-06-1999
			AU	653582 B2	06-10-1994
			CA	2077662 A1	09-07-1992
			DE	69210689 D1	20-06-1996
			DE	69210689 T2	21-11-1996
			EP	0520068 A1	30-12-1992
			KR	228688 B1	01-11-1999
			AT	144364 T	15-11-1996
			AT	138238 T	15-06-1996
			AU	649786 B2	02-06-1994
			AU	1194292 A	17-08-1992
			AU	1199792 A	17-08-1992
			CA	2077668 A1	09-07-1992
			DE	69214523 D1	21-11-1996
			DE	69214523 T2	27-03-1997
			DK	519055 T3	24-03-1997
			DK	520068 T3	15-07-1996
			EP	0519055 A1	23-12-1992
			ES	2093250 T3	16-12-1996
			ES	2087522 T3	16-07-1996
			JP	3197012 B2	13-08-2001
			JP	5505504 T	12-08-1993
			JP	5505298 T	05-08-1993
			JP	2003099099 A	04-04-2003
			KR	228687 B1	01-11-1999
			SG	49883 A1	15-06-1998
			SG	49884 A1	15-06-1998
			WO	9212607 A1	23-07-1992
			WO	9212608 A1	23-07-1992
			US	5400433 A	21-03-1995
			US	5608805 A	04-03-1997
			US	5274740 A	28-12-1993
			US	5390256 A	14-02-1995
US 5703999	A	30-12-1997	DE	4217276 C1	08-04-1993
			AT	142829 T	15-09-1996
			AU	683614 B2	20-11-1997
			AU	4059393 A	30-12-1993
			CA	2118402 A1	09-12-1993
			WO	9325015 A1	09-12-1993
			DE	59303771 D1	17-10-1996
			EP	0642719 A1	15-03-1995
			JP	8501666 T	20-02-1996

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

13

ANNEX TO THE EUROPEAN SEARCH REPORT
ON EUROPEAN PATENT APPLICATION NO.

EP 03 25 3970

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

18-09-2003

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
US 5703999 A		KR NO RU	240439 B1 944285 A 2119259 C1	15-01-2000 10-11-1994 20-09-1998

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82